

SEQUENCE LISTING

Chanc.
Rev/
~~5~~

(1) GENERAL INFORMATION:

(i) APPLICANT:

- (A) NAME: NOVARTIS AG
- (B) STREET: Schwarzwaldallee 215
- (C) CITY: Basel
- (E) COUNTRY: Switzerland
- (F) POSTAL CODE (ZIP): 4058
- (G) TELEPHONE: +41 61 69 11 11
- (H) TELEFAX: + 41 61 696 79 76
- (I) TELEX: 962 991

(ii) TITLE OF INVENTION: Improvements in or relating to organic compounds

(iii) NUMBER OF SEQUENCES: 33

(iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6695 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Daucus carota

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 3696..6617

(ix) FEATURE:

(A) NAME/KEY: intron

(B) LOCATION: 3731..3802

(ix) FEATURE:

(A) NAME/KEY: intron

(B) LOCATION: 3851..3979

(ix) FEATURE:

(A) NAME/KEY: intron

(B) LOCATION: 4124..4211

(ix) FEATURE:

(A) NAME/KEY: intron

(B) LOCATION: 4284..4357

(ix) FEATURE:

(A) NAME/KEY: intron

(B) LOCATION: 4430..4528

(ix) FEATURE:

(A) NAME/KEY: intron

(B) LOCATION: 4642..4757

(ix) FEATURE:

- (A) NAME/KEY: intron
- (B) LOCATION: 4890..4967

(ix) FEATURE:

- (A) NAME/KEY: intron
- (B) LOCATION: 5295..5803

(ix) FEATURE:

- (A) NAME/KEY: intron
- (B) LOCATION: 6197..6339

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

TCTAGATGAC GAAATCGCGC TACCTTGAT TINGAAATAC TAGGTTGTAG TATCTTGATT	60
AGTTTTTGG ATATCTTGCT GTAATTCTT TAGGAGATGC AAACGGTCCT CATTAAATAT	120
GAGCCCTTGT GACTTGACAA AAGTATCTAG CATGTTGAT CACGAGGTAG CTAAAAAGTA	180
GCGTGTGTTGA TTAAGCACAT AATATTGTAT TGGGCCTATT GGCTATCAAT GAAGTTTGAT	240
GCAAGTATAT AGCTTGTATT ATGCATGTGA TGAGGGTATA TAAAAGAAGT AAAGAACATT	300
CTCTCGTAGC ATTCACTTTT CTCTTGCTA TAGTTAACGA GTTTTGTAC ACATGACGTT	360
GAAACTGGAT GTGTCTGTC TTCCATCTAA GTTTGGATTA CCTGATAGAT GCTCAACTTC	420
TTCGTCAGCC TTTCTTTCG GATTTTCCC AAGACAAGAT TCTTTAGTTA ATAGTTATTG	480
CCTCTGGTGGC TTGTGTGCAT TTTAGGAATC TTACTCTGTT TTTTAATGGA GAAACGAAAC	540
CTACCTTTTT TCTCTGTC CCCCCATGTA TATCACCTGC TTGGAGGGGT TTAGACTTTA	600
TCCACCTAAA CTATTCATGT TTACCAGACA AGCTATACGT TTTATCCCCC CCCCCCGCGG	660
ACCTGNGGAC AAAAGAAGCG CTGATGAACT GATTAAATCC GTGTTTATT ATATTACACA	720

TTGATGCTTC ATGGAGCTAA TATCTTGGT TAAATTTCAT GTATATATAT ACCCTCCCT 780
CTTGTGATGG CAGTGGCCCC TCGTTAATT AGCGTACTTA ATTATCTGAT GGATACTGTA 840
TGCTGGCAG ATGATGTCAT CAGATTATAC CATTGGTGT GCTCTACAAA ATAAAAAAC 900
TCTATTTATG TTCATCTTTT TGGTAACAAG TAACTAATTG ATGCGCTATG TTGACAGGCG 960
ATGCATTACA CAACTTACGA ACTAGCTTGC AAGATCCAA CAATGTCCTG CAGAGCTGGG 1020
ATCCAACCCT TGTGAACCCCT TGCACATGGT TTCATGTGAC ATGTAACAAT GAAAACAGTG 1080
TTATAAGAGT GTAGGTCACT TCCCTTATTA ATTTTTTAG CAAGTTACGA ATATTTACTC 1140
AATTGAGCAG ATGTCCTTT AAATATTTT CTTTAATTTC TTAGCTAACG GGAGCATCTA 1200
TCTTAAGTAT CTCTACTGAA TTTAAGACAT AATACATTTC TTTAAAAAAAT CTATTAGAGT 1260
GTTTTTCCG CACAGCGCAC ATATATCTTT TTTCGGTAA TTCAGACAAC CTTTCTCCCG 1320
ACGATAAAAT AATATAAGAT TAACTCCTTG AACTAATTTC TTATTTCTC TTTCTTTTA 1380
TGTTCCTTGC AGAAAGTTTC TTATGGCTT TTGIGAAAAG TACATTCTAT GATAATTTC 1440
TGGCAACTCA TATAAATTAA TATATATTCC ATGTAGTTAT AAGTTAAAAA AAGCTTCCTA 1500
TTAATTCCAA GATAGAGGTT CATTTTATAA GTTGGGCAT CCATGAGTTT TTGAAAATGT 1560
CAGAAATTTC GTTGAGTTAA TTTTACCTAC CAACTTTAT GGCCTCATGC AGTGATCTTG 1620
GGAATGCAGC ATTATCTGGT CAATTGGTTC CTCTGGCCA GTTGAAAAT TTACAATACT 1680
TGTAAGACCA TATCACTTGG AATGCTTAG TTTTATACA GCACAATGCT TTCAATATCT 1740
GTTAAAAGTG TGAAAAGTT GACTTTCTAG CTTCAGCAGT TGTTGGATA ATATCTATGA 1800

AGCACTTAAA AGGCTGGCA ATTTTTTGT TATTATTCA AATATTGTTA ATTGTTACTA 1860
CTTAATATGA TAAACTGATT TAACTCCTCA TGATTGGCT CAGTCAAATG TGCCCTCATT 1920
AGTCACATNA TAAAATTGGN GGGTTGGACA AATATAACIT CTTTTCTTAA GGTCCAGAAA 1980
GAGCACTTAT CAACCTTGTC TAGCGCATAA CGTCACAGTG GGTCAAGTCAC GGGCTATCCA 2040
GTTTGGGGAG GTTTAAATGA GCACITATT ACCTTGTCTT TTAAACGTCT GAGGATGTTA 2100
TTAAAGTCTG CATCATTCAG AGTTTAAATT AGCACTTICA GTTGTATTAT GAATGGTACA 2160
TGAAAGATAC ATATCTTAAT GTTCTATGC CTGTTCAAC ATGTCCTAA TATTCTGTTA 2220
TCTTGTCTAT CTTAAAAATG GCACGTGATTA AAATGTGAGA AAGGTAGTCT TCCAATACCA 2280
TTTCATGTAT ACCAGAGAAT ATCATAATT TTAAATCA TAAGTTGGC CCTAGAGTTT 2340
TCTCAGTATT GGTCTATTAA TATTTCCAC CATTAGAAC TGTTGTGTC GATGAAAATC 2400
TTGGACITCC ACAGAAGATC TTATAGTAA AGTATTCTT AGATCTGATG ATGAAAGTTG 2460
TCATGGTGTG GCCTGTCCC GAATTAAAT CAATCCATG TCACATGTT GTGATCTGA 2520
CTACTCACTG TTAATCGAAG AGTAACATT TGTGAATTAA ATGCTTTTT TTTGTCTT 2580
CATGCTTAGC GTTATAAAGG TCTACGTCTG ACTATGGTTT TTAACATGTT ATAGTTTGT 2640
ACTGACAAGT TTAAAGTTTC TCTTGTCTAC GAATTAAGAA TATATAATAT AAAACGCTTT 2700
AACTTTCTCT GTGGAAGGTG TTCTTACCTT TTTATATATA TATATAGATA CTCAGACTCT 2760
GCTGGCAATT ATATCTTACG AACTTACGAG TATACAGAAC TTGTATATTA GGTCAGATG 2820
AGTGGCTGTA GTAGAACACC TTAAGCAAGA ACTTAATCAT GAGGTTCAA CCTTTAACT 2880
TTCTTTTAG ATTTTTCAA GTTATGGAA AATGTACCT CATGATCGTG GTTCTTCC 2940

ATAAACTTTC CATATAAGTC CGTTTCTTGA CGTTTTCATG TAAGCTGTG ACGAGTGATT 3000
ATTAGCGGTT CTTTCAATAA TCATAATGTG TCTCACTTTG ATGAGGCCTG TACTTATTAT 3060
TGCACCTTGC ACTTTAACCTT GATCCTCATG TCATCTTGTAT TGTCATAGTC TACTAACCGA 3120
GTGAAACATG GTTTATCATG TCTTTGAGG TAACAATGTA GCTTTCACCT CTGTCCTTGA 3180
TATAGGTTTA AGGCTTGCAC CTCCCCTAG CCCTTGTG TTCTTATCAC AGTTCACACA 3240
CCTACTAGCA CTGTCACCT CTAGTCCTT GTCCGAAAT AGTAAGAAGT TTCTTTCGCA 3300
TAATAGTGGA TGATCATTAA AGAAATAGTG AATCAAATTAA TCGTGTATT GTGTTGTAC 3360
TTTGGAAATTAA AATGAGTTGC TGAACATTGT TGCTGTTAT CGTTGTCAAG GCTTGCCTAA 3420
GGAAGGCGAT TAGTAAGAGT GGGCATCCAA GCGCCTTTAT CTGAAAGGGG CGGGCGGCAC 3480
GTTGTGGATT CTGGGTGCT ATTAGAGGAC ATTATCTATA TATACTGATT ATTATTAGA 3540
ATATAAATCA ACTACTATAT TTTCTTGT AATGTTATA TAGAAATCCC ACTCGTAAAC 3600
TTGACAAATAA CCATTGAAAT ATTTGAACCT AATTAATTAG TAGTGTCAAG TTTAAATTCA 3660
AACTCATTAA ATTTCATTAA AAAAAATAAT TCTATATGAA TCGTAACAGT ATAAATATAT 3720
TAAATTACAT GTATGTGTGC CTATATATAG CTGAATGTCT AATAGACTCC AAGACGGCTG 3780
CTCTTACTGC CTAGGGGTCC AGGCAGTTCA CTGATGCTTA CCTTGACAAA TATGGGGTTC 3840
GTATGACATT GTTGGGGATC CCTATCACTG GATTCTGTG TTGCTGACCC TCTGTTCAAT 3900
TGATTTTCAT TGATGTAGTA TTACTAGTT TATAAATATT CTTTATTGCA ATAATTAAAC 3960
TGGAGTTAA CAATGACAGG GAGCTTACA GCAATAACAT AAGTGGACCA ATTCTAGTG 4020

ATCTTGGGAA TCTGACAAAT TTGGTGAGCT TGGACCTATA CATGAATAGC TTCTCTGGAC 4080
CTATACCGGA CACATTAGGA AAGCTTACAA GGCTAAGATT CTGTATGAC TACAAATCTT 4140
CACTAGTTT TAACTTAACG CAATTTGATT ATCCCTCAA GTGATTGATT ATATCACAAA 4200
TTACTGGATA GGCCTCTCAA CAACAACTGC CTCTCTGGTC CAATTCCAAT GTCACTGACT 4260
AATATTACAA CTCTCAAGT CCTGTAAGTA TTCCGACCTT TCCAGATAGT TTGTTGTTG 4320
TGGATGTTTC AATTTAATA CTAAATATGT TCATCAGGGA TTTATCAAAC AATCGGCTAT 4380
CAGGACCAGT ACCGGATAAT GGCTCATTTT CTGTTTAC ACCTATCAGG TTTAATGCTA 4440
GTAATATCTT TAATATTATG GTCTTACTT CTACTGCGAA AGCTATGATA ATATTTTTT 4500
TCTCCTTCAT ATATTATCAC TTTCGCAGTT TTGGCAATAA TTGAAATTA TGIGGACCTG 4560
TAACTGGGAG GCGCTGCCCT GGATCTCCCC CATTCTCTCC ACCACCTCCG TTCACTCCCAC 4620
CATCAACAGT ACAGCCTCCA GGTGATTTAG TTTTATAATT AATTCCCGTA ATTAATTATA 4680
TGACTGTAAA AATTGGTGTGTT AATTCACCA GTTGCAGATA AAGTATTTCCTCTC 4740
TTCTTATTAT TATGAAGGAC AAAATGGTCC CACTGGAGCT ATTGCTGGGG GAGTAGCTGC 4800
TGGTGCTGCT TTACTGTTG CTGCACCTGC AATGGCATTT GCATGGTGGC GGAGAAGAAA 4860
ACCGCGAGAA CATTCTTTG ATGTGCCAGG TTAGTCCTGT AAATAGATAT CTATTGAAGC 4920
GCTTACTGTC TGTGGACTTT GTTTCACTG TCATTAGTTA ACTTCAGCTG AAGAGGACCC 4980
AGAAGTGCAC CTGGTCAAC TGAAGAGGTT TTCTCTGCGA GAATTGCAAG TCGCAACGGA 5040
TACTTTAGT ACCATCCTTG GAAGAGGTGG ATTTGTAAG GTGTATAAGG GACGCCCTG 5100
TGATGGCTCA CTGTAGCAG TAAAGAAGAA CGAACACCAG GTGGCGAGCT 5160

GCAGTTCAA ACAGAAGTGG AAATGATTAG CATGGCTGTG CATCGAAATC TTCTGCGTCT 5220
ACGTGGTTTC TGCATGACAC CTACCGAGCG GCTTCCTGTA TATCCATACA TGGCTAATGG 5280
AAGTGTGCG TCATGTTAA GAGGTATCTC AGTTACAATT ACCATAACTT GCCAGAAGTT 5340
TGTTGATTA AAAATGAAAT ATAACCTCCCT ACACATATGTT AAGGTGTTAT AATTCTGAG 5400
CAGATCTTAT TTCCCATTGC AAGATACCAAG TTATTATTGT TTTTCTGTA ATTGATACCG 5460
GTTATATTC TTCTCTGAT TTGGTTATAT GCAAGGATTT CGAGTCTAAT AAGTTATCAA 5520
ACTGGATGCT ATGTTATTC TGCAATTGAA TTCTGCTTC ATGTGCCAAA ATATATATGA 5580
TTCACCTGG AATCATCTTA TAATATACTG TGAAAGTCA GCTGTTGACT TTCATCATTA 5640
ATTAGCTTC ATAAATCAGA ATCTGCCTAG TGACCTTTAC CGACATACTC TAAACCTTC 5700
TTATGGCCCT GTATATAATC GTCCCCACTTA CTTTATTCAAG TTGTCTGCT CTCTGAATT 5760
TTGATCTGTA CATTGTGATG TCTTGTCTTC ATCAAATGTA GAGCGTCAGC CATCAGAAC 5820
TCCCCCTGAT TGGCCAACTA GGGAGAGGAT TGCACTAGGA TCTTCTAGGG GCCTATCTAA 5880
ATTGCATGAC CATTGTGATC CCAAGATTAT CCATCGCGAT GTAAAAGCTG CAAATATATT 5940
ATTGGACGAA GAATTGAGG CTGTTGTAGG TGATTTGGG TTAGCTAGGC TCATGGATTA 6000
CAAGGATACC CATGTTACGA CTGCTGTAAAG GGGTACCAATT GGGCACATAG CTCCCGAGTA 6060
CCTCTCGACT GGAAAGTCAT CAGAGAAGAC CGATGTCCTT GGTTATGGGA TAATGCTCCT 6120
AGAGCTCAATT ACTGGACAGA GGGCTTTGA TCTTGCTCGC CTTGCGAACG ATGATGATGT 6180
TAATGTTGTTG GATTGGGTAT GTGTCCCGGG TGTTCTTGT GTTAATTATT TCACATATTA 6240

GTCCTTACTA CTTTGTGTC GCGCTTGTT TTTATTCCT GCCTGTATTG GATTCTTAGT	6300
CATGTTATGC ATATTGACCT GCTTGTCAAT GTCTTTAGG TTAAAAGCCT TTGAAAGAG	6360
AAAAAGTTGG AGATGCTGGT CGATCCTGAC CTGCAGAAC AATTACATTGA CACAGAAGTT	6420
GAGCAGCTTA TTCAAGTAGC ATTACTCTGT ACCCAGGGTT CGCCAATGGA GCGGCCTAAG	6480
ATGTCAGAGG TAGTCGAAT GCTTGAAGGT GATGCCCTTG CAGAAAAGTG GGACGAGTGG	6540
CAAAAAGTIG AAGTCATCCA TCAAGACGTA GAATTAGCTC CACATCGAAC TTCTGAATGG	6600
ATCCTAGACT CGACAGATAA CTTGCATGCT TTGAAATTAT CTGGTCCAAG ATAAACAGCA	6660
TATAAAATGT AATGAAATTA ATATTTTTA TGGTT	6695

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1815 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Daucus carota

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 94..1752

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

GACAAATACC ATTGAAATAT TTGAACCTAA TTAATTAGTA GIGTCAGGGT TAAATTCAAA	60
CTCATTAAAT TTTACTTTAA AAAATAATT C TAT ATG AAT CGT AAC AGT ATA AAT Met Asn Arg Asn Ser Ile Asn	114
1 5	
ATA TTA AAT TAC ATG CAG TTC ACT GAT GCT CTT GAC AAA TAT GGG	162
Ile Leu Asn Tyr Met Gln Phe Thr Asp Ala Tyr Leu Asp Lys Tyr Gly	
10 15 20	
GTT CTT ATG ACA TTG GAG CTT TAC AGC AAT AAC ATA AGT GGA CCA ATT	210
Val Leu Met Thr Leu Glu Leu Tyr Ser Asn Asn Ile Ser Gly Pro Ile	
25 30 35	
CCT AGT GAT CTT GGG AAT CTG ACA AAT TTG GTG AGC TTG GAC CTA TAC	258
Pro Ser Asp Leu Gly Asn Leu Thr Asn Leu Val Ser Leu Asp Leu Tyr	
40 45 50 55	
ATG AAT AGC TTC TCT GGA CCT ATA CCG GAC ACA TTA GGA AAG CTT ACA	306
Met Asn Ser Phe Ser Gly Pro Ile Pro Asp Thr Leu Gly Lys Leu Thr	
60 65 70	
AGG CTA AGA TTC TTG CGT CTC AAC AAC AAC AGC CTC TCT GGT CCA ATT	354
Arg Leu Arg Phe Leu Arg Leu Asn Asn Asn Ser Leu Ser Gly Pro Ile	
75 80 85	
CCA ATG TCA CTG ACT AAT ATT ACA ACT CTT CAA GTC CTG GAT TTA TCA	402
Pro Met Ser Leu Thr Asn Ile Thr Thr Leu Gln Val Leu Asp Leu Ser	
90 95 100	
AAC AAT CGG CTA TCA GGA CCA GTA CCG GAT AAT GGC TCA TTT TCT TTG	450
Asn Asn Arg Leu Ser Gly Pro Val Pro Asp Asn Gly Ser Phe Ser Leu	
105 110 115	
TTT ACA CCT ATC AGT TTT GCC AAT AAT TTG AAT TTA TGT GGA CCC GTA	498

Phe Thr Pro Ile Ser Phe Ala Asn Asn Leu Asn Leu Cys Gly Pro Val
120 125 130 135

ACT GGG AGG CCC TGC CCT GGA TCT CCC CCA TTT TCG CCA CCA CCT CCG 546
Thr Gly Arg Pro Cys Pro Gly Ser Pro Pro Phe Ser Pro Pro Pro Pro
140 145 150

TTC ATC CCA CCA TCA ACA GTA CAG CCT CCA GGA CAA AAT GGT CCC ACT 594
Phe Ile Pro Pro Ser Thr Val Gln Pro Pro Gly Gln Asn Gly Pro Thr
155 160 165

GGA GCT ATT GCT GGG GGA GTA GCT GCT GGT GCT GCT TTA CTG TTT GCT 642
Gly Ala Ile Ala Gly Gly Val Ala Ala Gly Ala Ala Leu Leu Phe Ala
170 175 180

GCA CCT GCA ATG GCA TTT GCA TGG TGG CGG AGA AGA AAA CCG CGA GAA 690
Ala Pro Ala Met Ala Phe Ala Trp Trp Arg Arg Arg Lys Pro Arg Glu
185 190 195

CAT TTC TTT GAT GTG CCA GCT GAA GAG GAC CCA GAA GTG CAC CTT GGT 738
His Phe Phe Asp Val Pro Ala Glu Glu Asp Pro Glu Val His Leu Gly
200 205 210 215

CAA CTG AAG AGG TTT TCT CTG CGA GAA TTG CAA GTC GCA ACG GAT ACT 786
Gln Leu Lys Arg Phe Ser Leu Arg Glu Leu Gln Val Ala Thr Asp Thr
220 225 230

TTT AGT ACC ATA CTT GGA AGA GGT GGA TTT GGT AAG GTG TAT AAG GGA 834
Phe Ser Thr Ile Leu Gly Arg Gly Phe Gly Lys Val Tyr Lys Gly
235 240 245

CGC CTT GCT GAT GGC TCA CTT GTA GCA GTT AAA AGG CTT AAA GAA GAA 882
Arg Leu Ala Asp Gly Ser Leu Val Ala Val Lys Arg Leu Lys Glu Glu
250 255 260

CGA ACA CCA GGT GGT GAG CTG CAG TTT CAA ACA GAG GTG GAA ATG ATT 930
Arg Thr Pro Gly Gly Glu Leu Gln Phe Gln Thr Glu Val Glu Met Ile

265

270

275

AGC ATG GCT GTG CAT CGA AAT CTT CTG CGT CTA CGT GGT TTC TGC ATG Ser Met Ala Val His Arg Asn Leu Leu Arg Leu Arg Gly Phe Cys Met	280	285	290	295	978
ACA CCA ACA GAG CGG CTT CTT GTA TAT CCA TAC ATG GCT AAT GGA AGT Thr Pro Thr Glu Arg Leu Leu Val Tyr Pro Tyr Met Ala Asn Gly Ser	300	305	310		1026
GTT GCG TCG TGT TTA AGA GAG CGT CAG CCA TCA GAA CCT CCC CTT GAT Val Ala Ser Cys Leu Arg Glu Arg Gln Pro Ser Glu Pro Pro Leu Asp	315	320	325		1074
TGG CCA ACT AGG AAG AGG ATT GCA CTA GGA TCT GCT AGG GGG CTT TCT Trp Pro Thr Arg Lys Arg Ile Ala Leu Gly Ser Ala Arg Gly Leu Ser	330	335	340		1122
TAT TTG CAT GAC CAT TGT GAT CCC AAG ATT ATC CAT CGT GAT GTA AAA Tyr Leu His Asp His Cys Asp Pro Lys Ile Ile His Arg Asp Val Lys	345	350	355		1170
GCT GCA AAT ATA TTA TTG GAC GAA GAA TTT GAG GCT GTT GTA GGT GAT Ala Ala Asn Ile Leu Leu Asp Glu Glu Phe Glu Ala Val Val Gly Asp	360	365	370	375	1218
TTT GGG TTA GCT AGG CTC ATG GAT TAC AAG GAT ACC CAT GTT ACA ACT Phe Gly Leu Ala Arg Leu Met Asp Tyr Lys Asp Thr His Val Thr Thr	380	385	390		1266
GCT GTA AGG GGT ACC TTG GGC TAC ATA GCT CCC GAG TAC CTC TCG ACT Ala Val Arg Gly Thr Leu Gly Tyr Ile Ala Pro Glu Tyr Leu Ser Thr	395	400	405		1314
GGA AAG TCA TCA GAG AAG ACC GAT GTC TTT GGT TAT GGG ATT ATG CTC Gly Lys Ser Ser Glu Lys Thr Asp Val Phe Gly Tyr Gly Ile Met Leu	410	415	420		1362

TTA GAG CTC ATT ACT GGA CAG AGA GCT TTT GAT CTT GCT CGC CTT GCG. 1410
Leu Glu Leu Ile Thr Gly Gln Arg Ala Phe Asp Leu Ala Arg Leu Ala
425 430 435

AAC GAT GAT GAT GTT ATG TTG TTG GAT TGG GTT AAA AGC CTT TTG AAA 1458
Asn Asp Asp Asp Val Met Leu Leu Asp Trp Val Lys Ser Leu Leu Lys
440 445 450 455

GAG AAA AAG TTG GAG ATG CTG GTC GAT CCT GAC CTG GAG AAC AAT TAC 1506
Glu Lys Lys Leu Glu Met Leu Val Asp Pro Asp Leu Glu Asn Asn Tyr
460 465 470

ATT GAC ACA GAA GTT GAG CAG CTT ATT CAA GTA GCA TTA CTC TGT ACC 1554
Ile Asp Thr Glu Val Glu Gln Leu Ile Gln Val Ala Leu Leu Cys Thr
475 480 485

CAG GGT TCG CCA ATG GAG CGG CCT AAG ATG TCA GAG GTA GTC CGA ATG 1602
Gln Gly Ser Pro Met Glu Arg Pro Lys Met Ser Glu Val Val Arg Met
490 495 500

CTT GAA GGT GAT GGC CTT GCA GAA AAG TGG GAC GAG TGG CAA AAA GTA 1650
Leu Glu Gly Asp Gly Leu Ala Glu Lys Trp Asp Glu Trp Gln Lys Val
505 510 515

GAA GTC ATC CAT CAA GAC GTA GAA TTA GCT CCA CAT CGA ACT TCT GAA 1698
Glu Val Ile His Gln Asp Val Glu Leu Ala Pro His Arg Thr Ser Glu
520 525 530 535

TGG ATC CTA GAC TCG ACA GAT AAC TTG CAT GCT TTT GAA TTA TCT GGT 1746
Trp Ile Leu Asp Ser Thr Asp Asn Leu His Ala Phe Glu Leu Ser Gly
540 545 550

CCA AGA TAAACAGCAT ATAAAATGTG AATGAAATTAA ATATTTTTA TGGTTAAAAA 1802
Pro Arg

AAAAAAAAAA AAA

1815

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 553 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Met Asn Arg Asn Ser Ile Asn Ile Leu Asn Tyr Met Gln Phe Thr Asp
1 5 10 15

Ala Tyr Leu Asp Lys Tyr Gly Val Leu Met Thr Leu Glu Leu Tyr Ser
20 25 30

Asn Asn Ile Ser Gly Pro Ile Pro Ser Asp Leu Gly Asn Leu Thr Asn
35 40 45

Leu Val Ser Leu Asp Leu Tyr Met Asn Ser Phe Ser Gly Pro Ile Pro
50 55 60

Asp Thr Leu Gly Lys Leu Thr Arg Leu Arg Phe Leu Arg Leu Asn Asn
65 70 75 80

Asn Ser Leu Ser Gly Pro Ile Pro Met Ser Leu Thr Asn Ile Thr Thr
85 90 95

Leu Gln Val Leu Asp Leu Ser Asn Asn Arg Leu Ser Gly Pro Val Pro
100 105 110

Asp Asn Gly Ser Phe Ser Leu Phe Thr Pro Ile Ser Phe Ala Asn Asn
115 120 125

Leu Asn Leu Cys Gly Pro Val Thr Gly Arg Pro Cys Pro Gly Ser Pro
130 135 140

Pro Phe Ser Pro Pro Pro Phe Ile Pro Pro Ser Thr Val Gln Pro
145 150 155 160

Pro Gly Gln Asn Gly Pro Thr Gly Ala Ile Ala Gly Gly Val Ala Ala
165 170 175

Gly Ala Ala Leu Leu Phe Ala Ala Pro Ala Met Ala Phe Ala Trp Trp
180 185 190

Arg Arg Arg Lys Pro Arg Glu His Phe Phe Asp Val Pro Ala Glu Glu
195 200 205

Asp Pro Glu Val His Leu Gly Gln Leu Lys Arg Phe Ser Leu Arg Glu
210 215 220

Leu Gln Val Ala Thr Asp Thr Phe Ser Thr Ile Leu Gly Arg Gly Gly
225 230 235 240

Phe Gly Lys Val Tyr Lys Gly Arg Leu Ala Asp Gly Ser Leu Val Ala
245 250 255

Val Lys Arg Leu Lys Glu Glu Arg Thr Pro Gly Gly Glu Leu Gln Phe
260 265 270

Gln Thr Glu Val Glu Met Ile Ser Met Ala Val His Arg Asn Leu Leu
275 280 285

Arg Leu Arg Gly Phe Cys Met Thr Pro Thr Glu Arg Leu Leu Val Tyr
290 295 300

Pro Tyr Met Ala Asn Gly Ser Val Ala Ser Cys Leu Arg Glu Arg Gln
305 310 315 320

Pro Ser Glu Pro Pro Leu Asp Trp Pro Thr Arg Lys Arg Ile Ala Leu

B63474 20080709

325

330

335

Gly Ser Ala Arg Gly Leu Ser Tyr Leu His Asp His Cys Asp Pro Lys
340 345 350

Ile Ile His Arg Asp Val Lys Ala Ala Asn Ile Leu Leu Asp Glu Glu
355 360 365

Phe Glu Ala Val Val Gly Asp Phe Gly Leu Ala Arg Leu Met Asp Tyr
370 375 380

Lys Asp Thr His Val Thr Thr Ala Val Arg Gly Thr Leu Gly Tyr Ile
385 390 395 400

Ala Pro Glu Tyr Leu Ser Thr Gly Lys Ser Ser Glu Lys Thr Asp Val
405 410 415

Phe Gly Tyr Gly Ile Met Leu Leu Glu Leu Ile Thr Gly Gln Arg Ala
420 425 430

Phe Asp Leu Ala Arg Leu Ala Asn Asp Asp Asp Val Met Leu Leu Asp
435 440 445

Trp Val Lys Ser Leu Leu Lys Glu Lys Lys Leu Glu Met Leu Val Asp
450 455 460

Pro Asp Leu Glu Asn Asn Tyr Ile Asp Thr Glu Val Glu Gln Leu Ile
465 470 475 480

Gln Val Ala Leu Leu Cys Thr Gln Gly Ser Pro Met Glu Arg Pro Lys
485 490 495

Met Ser Glu Val Val Arg Met Leu Glu Gly Asp Gly Leu Ala Glu Lys
500 505 510

Trp Asp Glu Trp Gln Lys Val Glu Val Ile His Gln Asp Val Glu Leu
515 520 525

Ala Pro His Arg Thr Ser Glu Trp Ile Leu Asp Ser Thr Asp Asn Leu
530 535 540

His Ala Phe Glu Leu Ser Gly Pro Arg
545 550

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

TTTTTTTTT TGC

13

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GGGATCTAAG

10

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

ACACGTGGTC

10

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

TCAGCACAGG

10

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

TTTTTTTTT TCTG

14

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

TTTTTTTTT TCA

13

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

GACATCGTCC

10

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

CCCTACTGGT

10

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

ACACGTGGTC

10

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

GGTGACTGTC

10

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

TCTTGGACCA GATAATTC

18

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

CTCTGATGAC TTTCCAGTC

19

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

AATGGCATT GCATGG

16

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 amino acids
- (B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Daucus carota

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

Ser Pro Pro Pro Pro

1

5

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 8 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Daucus carota

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

His Arg Asp Val Lys Ala Ala Asn

S63477 00200750

1

5

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Daucus carota

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

Gly Thr Leu Gly Tyr Ile Ala Pro Glu

1

5

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4081 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

2025 RELEASE UNDER E.O. 14176

(A) ORGANISM: *Arabidopsis thaliana*

(vii) IMMEDIATE SOURCE:

(B) CLONE: *Arabidopsis SERK* gene

(ix) FEATURE:

(A) NAME/KEY: exon

(B) LOCATION: 1280..1367

(ix) FEATURE:

(A) NAME/KEY: exon

(B) LOCATION: 1796..1928

(ix) FEATURE:

(A) NAME/KEY: exon

(B) LOCATION: 2014..2085

(ix) FEATURE:

(A) NAME/KEY: exon

(B) LOCATION: 2203..2346

(ix) FEATURE:

(A) NAME/KEY: exon

(B) LOCATION: 2450..2521

(ix) FEATURE:

(A) NAME/KEY: exon

(B) LOCATION: 2617..2688

(ix) FEATURE:

(A) NAME/KEY: exon

(B) LOCATION: 2772..2884

(ix) FEATURE:

(A) NAME/KEY: exon

(B) LOCATION: 3015..3146

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 3305..3646

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 3760..4081

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

TCTAGAAACC TTTTGATCAT AATGAAAATA AAGAGTCAT CCACCACATG GGGTAAGCAT 60
AATGTGTGAT ATTTAAAGGG TAACAAATGT AATCTGCTTT TTATTTTACT TTTTACCTCT 120
ACTCAAATTG TATGGCAGT TTTTTTTTTT TTTTAAATGA TAAGACAAGT ATCTGTTAA 180
TGGTATTGTG ATGAAACAGT AGTAAAGTCA TATCGGGCAC GCCATACTAC TTCCACAGTG 240
GAACTTGGCC AAATTTGTC TTTGCCGTCT CTACAGTTTC TTCCACCAAA TTTTTGTG 300
ACAAAAACTCA AATCTTCAA TCTCATCTCT GCCAAAGTTG GGTTTAGAAA GAATATCAGC 360
AACACATAAT ATCTTTATTG TTGCATGGTT TATCAATCAC AAAATTCAAA ACCATTGTAA 420
AAAAAAATTC ACATTTTGG TATGAGATTG CTCACATGAT AGTGAACCTC TTTAACATTT 480
TAACTTTACT TTCATAAATA CGGGATTACG AATCTTACTT GCATTAAGAAA TTTAGAAAAG 540
GTTTTCTAC TAAAGAAAA AAGGGACCCA ACAGAGAGAG GTTGACCCAG GAGAAACGGG 600
TGCATAGCCT TAAGAGCTTT CAACTACTTT ACCCCAAACC CAAAGCGATG TCACTTCAA 660
CCATCTCTTC TCTCCCCCGA ACCCGTTTT TTGACCGGTC AGTTGGGCA GCAGCACCGT 720
TACGGGCAGC TTATATTCT CGTCCTCTC CTCACACCA CTGCATGCCA ATAATAAAG 780

CCCGTTGAGA TCTTTAAAAA TATTAATAA TATATCAACG AAAAAGCTAT TTTATTCTATA 840
AGAAGAAAAA GAGAGGAACA ACAACAACAC ACTAAATCATA GTTCTCTGG CAGGCTTGT 900
GTTCGGCTT AATAAAAAGC TCTTTGTAA TTATTACTTC ACGTAGATT TCCCCAAAAA 960
GCTCTTATT TTTGTAA AAAAAAAAGT TTCATCTTA TCACACTTTT GTTTACAGT 1020
GTGTGTGTA GAGAGAGAGT GTGGTTTGAT TGAGGAAAGA CGACGACGAG AACGCCGGAG 1080
AATTAGGATT TTTATTAT TTTTACTCT TTGTTGTT TAATGCTAAT GGGTTTTAA 1140
AAGGGTTATC GAAAAATGA GTGAGTTGT GTTGAGGTIG TCTCTGAAA GTGTTAATGG 1200
TGGTGATTT CGGAAGTTAG GGTTTCTCG GATCTGAAGA GATCAAATCA AGATTCGAAA 1260
TTTACCTTCG AATCATTCAAC TGTGGCTGC TTCTGCTAAT TTGGAAGGTT CGTGGTTACT 1320
CAATTACTCA GCTTTACTCG TTCTCAATT ACTTCTCGA TCTTTTTAA TTGGAGGTG 1380
AATCGCTATC TTAGTGTCT GCATTTGAT TTATGAAAAT TGTGTTGTT CTTGTATTT 1440
GTAAGATTTA GTGGCTAGTA CTTGAATAC ACTGTTTGC TTTCTTGT CAGATCAACT 1500
TTGTATATTG TAAAGGCATG TTCTTGGGT TGAAAAGCTG GGTTATTGTA TATCTTAAGA 1560
TTGATGTTGT TGATCCAAAC ATTCTCTGAA AGACTTCATT TGTTTTGGT TTTGTAAAGA 1620
ATTGTTAA TTATAGCCT CTAATCTCAG AGAGGCCIGT TTGAATAGTT CTCTCTGAA 1680
ATTAGACTTT TCACCAATTG ATGCTAATTG TGTAGATTG TTGTTCTGT TATAGGTGAT 1740
GCTTGCATA CTTGAGGGT TACTCTAGTT GATCCAAACA ATGCTTGCAG GAGCTGGGAT 1800
CCTACGCTAG TGAATCTTG CACATGGTTC CATGTCACCT GCAACAAACGA GAACAGTGT 1860
CCTACGCTAG TGAATCTTG CACATGGTTC CATGTCACCT GCAACAAACGA GAACAGTGT 1920

ATAAGAGTGT AAAGCTTTCT TCTACTAATC CCACCTTTTA AACCTTGACC TCAGCGTGGT 1980
TACCGACATT TTTGTTTCCTT TIGTCAAATA CAGTGATTIG GGGAAATGCAG AGTTATCTGG 2040
CCATTIAGTT CCAGAGCTTG GTGTGCTCAA GAATTIGCAG TATTIGTAAG TTCCACTTAT 2100
GCATCATGCT TTAACAAAAC AAATCCAAGA TTTGACAGAA GAAGCACTGG AGTTACCTTT 2160
TGTAAATTGAA ATCTTTTAA CAAGTTCTT ATTTCCTTAC AGGGAGCTTT ACAGTAACAA 2220
CATAACTGGC CCGATTCTTA GEAATCTTGG AAATCTGACA AACCTTAGTGA GTTGGATCT 2280
TTACTTAAAC AGCTTCTCOG GTCCTATTCC GGAATCATTG GGAAAGCTTT CAAAGCTGAG 2340
ATTTCCTGTGA GTATACATAT GCTTTACCGG CTCAGTTACA GTCCTTGTTT AATCTTAGGT 2400
TTTGTCCAA TTTTGTACTC TTTGCTGAAA ATTTACATG CAAGAATAGC CGGCTTAACA 2460
ACAACAGTCT CACTGGTCA ATTCTATGT CACTGACCAA TATTACTACC CTCAAGTGT 2520
TGTGAGTCCT CTCATTAACCTT TTCATTTATG TCTACTTCAT TCTCCCTCAG TTGATTGTT 2580
GAGTTAACGC ACTTAAACCTT GATGGATGCC ACACAGAGAT CTATCAAATA ACAGACTCTC 2640
TGGTTCAAGT CCTGACAATG GCTCCTTCAC ACTCTTCACA CCCATCAGGT TCTATGATT 2700
ATCCCTCTCA GTTATTTCAG TTGTTGTGTC AGTGTCTGAA CTTATTCTGA AACTTTCAATT 2760
TCCCTGTGCA GTTTGCTAA TAACCTAGAC CTATGTGGAC CTGTTACAAG TCACCCATGT 2820
CCTGGATCTC CCCCCGTTTC TCCCTCACCA CCTTTTATTC AACCTCCCCC AGTTTCCACC 2880
CCGAGTAAGC CTCCCTTTT TAGTTTACAT TATAGGAAAC AGAAGATGAA ATCTTTGCTT 2940
CTCTGTCAAT CCTTTTCTC ATATAACTCA TCTTGCCAAT AAGGCAATAA CCAAATGATC 3000

TAATTGATT TCAGGTGGGT ATGGTATAAC TGGAGCAATA GCTGGTGGAG TTGCTGCAGG	3060
TGCTGCTTG CTCTTGCTG CTOCTGAAT AGCCTTGTCT TGGTGGCGAC GAAGAAAGCC	3120
ACTAGATATT TTCTTCGATG TGCTGGTGA GTTTATTATT CGCATTAGTT TCTGTTCTTA	3180
GCCAGCAATT TIGTTTGCA GAAAAGTATT GGAACAAC TG TTAATGAAAA TCAATACATA	3240
AGTCATTGTT TTTAAGTTA CAAACTCTTT TGAGTAAAAT CTGATTGCA AAATCTCTAT	3300
GCAGCCGAAG AAGATCCAGA AGTCATCTG GGACAGCTCA AGAGGTTTC TTTGCCGGAG	3360
CTACAAGTGG CGAGTGATGG GTTTAGTAAAC AAGAACATTG TGGGCAGAGG TGGGTTGGG	3420
AAAGTCTACA AGGGACGCTT GGCAGACGGA ACTCTTGTIG CTGTCAAGAG ACTGAAGGAA	3480
GAGCGAACTC CAGGTGGAGA GCTCCAGTT CAAACAGAAG TAGAGATGAT AAGTATGGCA	3540
GTCATCGAA ACCTGTTGAG ATTACGAGGT TTCCTGTATGA CACCGACCGA GAGATTGCTT	3600
GTGTATCCTT ACATGGCAA TGGAAAGTGT GCTTCGTGTC TCAGAGGTAA AAACTAAACA	3660
ATTAAACATC TTGTGCTCTC TCTCAATTAC TTTGACGTGA AGTGTGTTT CATGTTTCC	3720
TTTATGGTT CATAATTGTT GGTTACACTA ATGACACAGA GAGGCCACCG TCACAACCTC	3780
CGCTTGATTG GCCAACGCGG AAGAGAATCG CGCTAGGCTC AGCTCGAGGT TTGCTTACC	3840
TACATGATCA CTGCGATCCG AAGATCATTC ACCGTGACGT AAAAGCAGCA AACATCCTCT	3900
TAGACGAAGA ATTGGAAGCG GTTGTGGAG ATTTCGGTT GGCAAAGCTA ATGGACTATA	3960
AAGACACTCA CGTGACAACA GCAGTCGGTG GCACCATCGG TCACATCGT CCAGAATATC	4020
TCTCAACCGG AAAATCTCTCA GAGAAAACCG ACGTTTCCG ATACGGAATC ATGCTTCTAG	4080
A	4081

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 494 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: N-terminal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

Met Glu Ser Ser Tyr Val Val Phe Ile Leu Leu Ser Leu Ile Leu Leu
1 5 10 15

Pro Asn His Ser Leu Trp Leu Ala Ser Ala Asn Leu Glu Gly Asp Ala
20 25 30

Leu His Thr Leu Arg Val Thr Leu Val Asp Pro Asn Asn Val Leu Gln
35 40 45

Ser Trp Asp Pro Thr Leu Val Asn Pro Cys Thr Trp Phe His Val Thr
50 55 60

Cys Asn Asn Glu Asn Ser Val Ile Arg Val Asp Leu Gly Asn Ala Glu
65 70 75 80

Leu Ser Gly His Leu Val Pro Glu Leu Gly Val Leu Lys Asn Leu Gln
85 90 95

Glu Leu Tyr Ser Asn Asn Ile Thr Gly Pro Ile Pro Ser Asn Leu Gly
100 105 110

Asn Leu Thr Asn Leu Val Ser Leu Asp Leu Tyr Leu Asn Ser Phe Ser
115 120 125

Gly Pro Ile Pro Glu Ser Leu Gly Lys Leu Ser Lys Leu Arg Phe Leu
130 135 140

Arg Leu Asn Asn Asn Ser Leu Thr Gly Ser Ile Pro Met Ser Leu Thr
145 150 155 160

Asn Ile Thr Thr Leu Gln Val Leu Asp Leu Ser Asn Asn Arg Leu Ser
165 170 175

Gly Ser Val Pro Asp Asn Gly Ser Phe Ser Leu Phe Thr Pro Ile Ser
180 185 190

Phe Ala Asn Asn Leu Asp Leu Cys Gly Pro Val Thr Ser His Pro Cys
195 200 205

Pro Gly Ser Pro Pro Phe Ser Pro Pro Pro Pro Phe Ile Gln Pro Pro
210 215 220

Pro Val Ser Thr Pro Ser Gly Tyr Gly Ile Thr Gly Ala Ile Ala Gly
225 230 235 240

Gly Val Ala Ala Gly Ala Ala Leu Leu Phe Ala Ala Pro Ala Ile Ala
245 250 255

Phe Ala Trp Trp Arg Arg Arg Lys Pro Leu Asp Ile Phe Phe Asp Val
260 265 270

Pro Ala Glu Glu Asp Pro Glu Val His Leu Gly Gln Leu Lys Arg Phe
275 280 285

Ser Leu Arg Glu Leu Gln Val Ala Ser Asp Gly Phe Ser Asn Lys Asn

290 295 300
Ile Leu Gly Arg Gly Gly Phe Gly Lys Val Tyr Lys Gly Arg Leu Ala
305 310 315 320
Asp Gly Thr Leu Val Ala Val Lys Arg Leu Lys Glu Glu Arg Thr Pro
325 330 335
Gly Gly Glu Leu Gln Phe Gln Thr Glu Val Glu Met Ile Ser Met Ala
340 345 350
Val His Arg Asn Leu Leu Arg Leu Arg Gly Phe Cys Met Thr Pro Thr
355 360 365
Glu Arg Leu Leu Val Tyr Pro Tyr Met Ala Asn Gly Ser Val Ala Ser
370 375 380
Cys Leu Arg Glu Arg Pro Pro Ser Gln Pro Pro Leu Asp Trp Pro Thr
385 390 395 400
Arg Lys Arg Ile Ala Leu Gly Ser Ala Arg Gly Leu Ser Tyr Leu His
405 410 415
Asp His Cys Asp Pro Lys Ile Ile His Arg Asp Val Lys Ala Ala Asn
420 425 430
Ile Leu Leu Asp Glu Glu Phe Glu Ala Val Val Gly Asp Phe Gly Leu
435 440 445
Ala Lys Leu Met Asp Tyr Lys Asp Thr His Val Thr Thr Ala Val Arg
450 455 460
Gly Thr Ile Gly His Ile Ala Pro Glu Tyr Leu Ser Thr Gly Lys Ser
465 470 475 480
Ser Glu Lys Thr Asp Val Phe Gly Tyr Gly Ile Met Leu Leu
485 490

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1106 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 142..795

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

TOGACCCACG CGTCGGTCCA ACTTCAATAA AGGGGAAACC AACGTAACCC TAATTTGCT	60
TTCTCCTCTT TGTTCAGAAA ATTTTCCCTT TACTCTCAAA TTCCTTTTCG ATTTCCTCT	120
CTTAAACCTC CGAAAGCTCA C ATG GCG TCT CGA AAC TAT CGG TGG GAG CTC	171
Met Ala Ser Arg Asn Tyr Arg Trp Glu Leu	
1 5 10	
TTC GCA GCT TCG TTA ACC CTA ACC TTA GCT TTG ATT CAC CTG GTC GAA	219
Phe Ala Ala Ser Leu Thr Leu Thr Leu Ala Leu Ile His Leu Val Glu	
15 20 25	
GCA AAC TCC GAA GGA GAT GCT CTC TAC GCT CTT CGC CGG AGT TTG ACA	267
Ala Asn Ser Glu Gly Asp Ala Leu Tyr Ala Leu Arg Arg Ser Leu Thr	
30 35 40	

GAT CCA GAC CAT GTC CTC CAG AGC TGG GAT CCA ACT CTT GTT AAT CCT Asp Pro Asp His Val Leu Gln Ser Trp Asp Pro Thr Leu Val Asn Pro	315
45 50 55	
TGT ACC TGG TTC CAT GTC ACC TGT AAC CAA GAC AAC CGC GTC ACT CGT Cys Thr Trp Phe His Val Thr Cys Asn Gln Asp Asn Arg Val Thr Arg	363
60 65 70	
GTG GAT TTG GGA AAT TCA AAC CTC TCT GGA CAT CTT GCG CCT GAG CTT Val Asp Leu Gly Asn Ser Asn Leu Ser Gly His Leu Ala Pro Glu Leu	411
75 80 85 90	
GGG AAG CTT GAA CAT TTA CAG TAT CTA GAG CTC TAC AAA AAC AAC ATC Gly Lys Leu Glu His Leu Gln Tyr Leu Glu Leu Tyr Lys Asn Asn Ile	459
95 100 105	
CAA GGA ACT ATA CCT TCC GAA CTT GGA AAT CTG AAG AAT CTC ATC AGC Gln Gly Thr Ile Pro Ser Glu Leu Gly Asn Leu Lys Asn Leu Ile Ser	507
110 115 120	
TTG GAT CTG TAC AAC AAC AAT CTT ACA GGG ATA GTT CCC ACT TTC TTG Leu Asp Leu Tyr Asn Asn Leu Thr Gly Ile Val Pro Thr Phe Leu	555
125 130 135	
GGA AAA TTG AAG TCT CTG GTC TTT TTA CGG CTT AAT GAC AAC CGA TTG Gly Lys Leu Lys Ser Leu Val Phe Leu Arg Leu Asn Asp Asn Arg Leu	603
140 145 150	
ACC GGT CCA ATC CTA GAG CAC TCA CGG CAA TCC CAA GCC TTT AAA GTT Thr Gly Pro Ile Leu Glu His Ser Arg Gln Ser Gln Ala Phe Lys Val	651
155 160 165 170	
GTT GAC GTC TCA AGC AAT GAT TTG TGT GGG ACA ATC CCA ACA AAC GGA Val Asp Val Ser Ser Asn Asp Leu Cys Gly Thr Ile Pro Thr Asn Gly	699
175 180 185	
CCC TTT GCT CAC ATT CCT TTA CAG AAC TTT GAG AAC AAC CCG AGA TTG	747

Pro Phe Ala His Ile Pro Leu Gln Asn Phe Glu Asn Asn Pro Arg Leu
190 195 200

GAG GGA CCG GAA TTA CTC GGT CTT GCA AGC TAC GAC ACT AAC TGC ACC 795
Glu Gly Pro Glu Leu Leu Gly Leu Ala Ser Tyr Asp Thr Asn Cys Thr
205 210 215

TGAAACAACT GGCAAAACCT GAAAATGAAG AATTGGGGGG TGACCTTGTA AGAACACTTC 855
ACCACITAT CAAATATCAC ATCTATTATG TAATAAGTAT ATATATGTAG TAAAAACAAA 915

AAAAATGAAG AATCGAATCG GTAATATCAT CTGGTCTCAA TTGAGAACCT CGAGGTCTGT 975
ATGTAAAATT TCTAAATGCG ATTTCGCTT ACTGTAATGT TCGGGTGTGG GATTCTGAGA 1035
AGTAACATT GTATTGGTAT GGTATCAAGT TGTTCTGCCT TGCTCTGAAA AAAAAAAAAA 1095
AAAAAAAAAA A 1106

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 218 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

Met Ala Ser Arg Asn Tyr Arg Trp Glu Leu Phe Ala Ala Ser Leu Thr
1 5 10 15

Leu Thr Leu Ala Leu Ile His Leu Val Glu Ala Asn Ser Glu Gly Asp
20 25 30

Ala Leu Tyr Ala Leu Arg Arg Ser Leu Thr Asp Pro Asp His Val Leu
35 40 45

Gln Ser Trp Asp Pro Thr Leu Val Asn Pro Cys Thr Trp Phe His Val
50 55 60

Thr Cys Asn Gln Asp Asn Arg Val Thr Arg Val Asp Leu Gly Asn Ser
65 70 75 80

Asn Leu Ser Gly His Leu Ala Pro Glu Leu Gly Lys Leu Glu His Leu
85 90 95

Gln Tyr Leu Glu Leu Tyr Lys Asn Asn Ile Gln Gly Thr Ile Pro Ser
100 105 110

Glu Leu Gly Asn Leu Lys Asn Leu Ile Ser Leu Asp Leu Tyr Asn Asn
115 120 125

Asn Leu Thr Gly Ile Val Pro Thr Phe Leu Gly Lys Leu Lys Ser Leu
130 135 140

Val Phe Leu Arg Leu Asn Asp Asn Arg Leu Thr Gly Pro Ile Leu Glu
145 150 155 160

His Ser Arg Gln Ser Gln Ala Phe Lys Val Val Asp Val Ser Ser Asn
165 170 175

Asp Leu Cys Gly Thr Ile Pro Thr Asn Gly Pro Phe Ala His Ile Pro
180 185 190

Leu Gln Asn Phe Glu Asn Asn Pro Arg Leu Glu Gly Pro Glu Leu Leu
195 200 205

Gly Leu Ala Ser Tyr Asp Thr Asn Cys Thr
210 215

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 981 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 104..757

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

AGTGTGAGTA ATTTAGTTTG CTTTCCTCCTC TTGGTCAGA AAATTTCCC TTTACTCTCA	60
AATTCCCTTT CGATTCCCT CTCTAAACC TCGAAAGCT CAC ATG GCG TCT CGA	115
Met Ala Ser Arg	
1	
AAC TAT CCG TGG GAG CTC TTC GCA GCT TCG TTA ACC CTA ACC TTA GCT	163
Asn Tyr Arg Trp Glu Leu Phe Ala Ala Ser Leu Thr Leu Thr Leu Ala	
5 10 15 20	
TTC ATT CAC CTG GTC GAA GCA AAC TCC GAA GGA GAT GCT CTC TAC GCT	211
Leu Ile His Leu Val Glu Ala Asn Ser Glu Gly Asp Ala Leu Tyr Ala	
25 30 35	
CTT CGC CGG AGT TTG ACA GAT CCA GAC CAT GTC CTC CAG AGC TGG GAT	259
Leu Arg Arg Ser Leu Thr Asp Pro Asp His Val Leu Gln Ser Trp Asp	
40 45 50	

CCA ACT CTT GTT AAT CCT TGT ACC TGG TTC CAT GTC ACC TGT AAC CAA 307
Pro Thr Leu Val Asn Pro Cys Thr Trp Phe His Val Thr Cys Asn Gln
55 60 65

GAC AAC CGC GTC ACT CGT GTG GAT TTG GGA AAT TCA AAC CTC TCT GGA 355
Asp Asn Arg Val Thr Arg Val Asp Leu Gly Asn Ser Asn Leu Ser Gly
70 75 80

CAT CTT GCG CCT GAG CTT GGG AAG CTT GAA CAT TTA CAG TAT CTA GAG 403
His Leu Ala Pro Glu Leu Gly Lys Leu Glu His Leu Gln Tyr Leu Glu
85 90 95 100

CTC TAC AAA AAC AAC ATC CAA GGA ACT ATA CCT TCC GAA CTT GGA AAT 451
Leu Tyr Lys Asn Asn Ile Gln Gly Thr Ile Pro Ser Glu Leu Gly Asn
105 110 115

CTG AAG AAT CTC ATC AGC TTG GAT CTG TAC AAC AAC AAT CTT ACA GGG 499
Leu Lys Asn Leu Ile Ser Leu Asp Leu Tyr Asn Asn Asn Leu Thr Gly
120 125 130

ATA GTT CCC ACT TCT TTG GGA AAA TTG AAG TCT CTG GTC TTT TTA CGG 547
Ile Val Pro Thr Ser Leu Gly Lys Leu Lys Ser Leu Val Phe Leu Arg
135 140 145

CTT AAT GAC AAC CGA TTG ACC GGT CCA ATC CCT AGA GCA CTC ACG GCA 595
Leu Asn Asp Asn Arg Leu Thr Gly Pro Ile Pro Arg Ala Leu Thr Ala
150 155 160

ATC CCA AGC CTT AAA GTT GTT GAC GTC TCA AGC AAT GAT TTG TGT GGA 643
Ile Pro Ser Leu Lys Val Val Asp Val Ser Ser Asn Asp Leu Cys Gly
165 170 175 180

ACA ATC CCA ACA AAC GGA CCC TTT GCT CAC ATT CCT TTA CAG AAC TTT 691
Thr Ile Pro Thr Asn Gly Pro Phe Ala His Ile Pro Leu Gln Asn Phe
185 190 195

GAG AAC AAC CCG AGA TTG GAG GGA CCG GAA TTA CTC GGT CTT GCA AGC 739

Glu Asn Asn Pro Arg Leu Glu Gly Pro Glu Leu Leu Gly Leu Ala Ser			
200	205	210	
TAC GAC ACT AAC TGC ACC TGAAACAACCT GGCAAAACCT GAAAATGAAG			787
Tyr Asp Thr Asn Cys Thr			
215			
AATTGGGGGG TGACCTTGTA AGAACACTTC ACCACTTTAT CAAATATCAC ATCTATTATG			847
TAATAAGTAT ATATATGTAG TAAAAACAAA AAAAATGAAG AATCGAATCG GTAATATCAT			907
CTGGTCTCAA TTGAGAACCTT CGAGGTCTGT ATGTAAAATT TCTAAATGCG ATTTCGCCT			967
AAATTACTCA CACT			981

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 218 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

Met Ala Ser Arg Asn Tyr Arg Trp Glu Leu Phe Ala Ala Ser Leu Thr

1 5 10 15

Leu Thr Leu Ala Leu Ile His Leu Val Glu Ala Asn Ser Glu Gly Asp

20 25 30

Ala Leu Tyr Ala Leu Arg Arg Ser Leu Thr Asp Pro Asp His Val Leu

35 40 45

Gln Ser Trp Asp Pro Thr Leu Val Asn Pro Cys Thr Trp Phe His Val

50

55

60

Thr Cys Asn Gln Asp Asn Arg Val Thr Arg Val Asp Leu Gly Asn Ser
65 70 75 80

Asn Leu Ser Gly His Leu Ala Pro Glu Leu Gly Lys Leu Glu His Leu
85 90 95

Gln Tyr Leu Glu Leu Tyr Lys Asn Asn Ile Gln Gly Thr Ile Pro Ser
100 105 110

Glu Leu Gly Asn Leu Lys Asn Leu Ile Ser Leu Asp Leu Tyr Asn Asn
115 120 125

Asn Leu Thr Gly Ile Val Pro Thr Ser Leu Gly Lys Leu Lys Ser Leu
130 135 140

Val Phe Leu Arg Leu Asn Asp Asn Arg Leu Thr Gly Pro Ile Pro Arg
145 150 155 160

Ala Leu Thr Ala Ile Pro Ser Leu Lys Val Val Asp Val Ser Ser Asn
165 170 175

Asp Leu Cys Gly Thr Ile Pro Thr Asn Gly Pro Phe Ala His Ile Pro
180 185 190

Leu Gln Asn Phe Glu Asn Asn Pro Arg Leu Glu Gly Pro Glu Leu Leu
195 200 205

Gly Leu Ala Ser Tyr Asp Thr Asn Cys Thr
210 215

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 789 base pairs
- (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 2..661

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

T CGA CCC ACG CGT CGG CGA AAC TAT CGG TGG GAG CTC TTC GCA GCT Arg Pro Thr Arg Pro Arg Asn Tyr Arg Trp Glu Leu Phe Ala Ala	46
1 5 10 15	
TCG TTA ATC CTA ACC TTA GCT TTG ATT CAC CTG GTC GAA GCA AAC TCC Ser Leu Ile Leu Thr Leu Ala Leu Ile His Leu Val Glu Ala Asn Ser	94
20 25 30	
GAA GGA GAT GCT CTT TAC GCT CTT CGC CGG AGT TTA ACA GAT CCG GAC Glu Gly Asp Ala Leu Tyr Ala Leu Arg Arg Ser Leu Thr Asp Pro Asp	142
35 40 45	
CAT GTT CTC CAG AGC TGG GAT CCA ACT CTT GTT AAT CCT TGT ACC TGG His Val Leu Gln Ser Trp Asp Pro Thr Leu Val Asn Pro Cys Thr Trp	190
50 55 60	
TTC CAT GTC ACC TGT AAC CAA GAC AAC CGC GTC ACT CGT GTG GAT TTG Phe His Val Thr Cys Asn Gln Asp Asn Arg Val Thr Arg Val Asp Leu	238
65 70 75	
GGG AAT TCA AAC CTC TCT GGA CAT CTT GCG CCT GAG CTT GGG AAG CTT Gly Asn Ser Asn Leu Ser Gly His Leu Ala Pro Glu Leu Gly Lys Leu	286

80	85	90	95	
GAA CAT TTA CAG TAT CTA GAG CTC TAC AAA AAC AAC AAC ATC CAA GGA ACT Glu His Leu Gln Tyr Leu Glu Leu Tyr Lys Asn Asn Ile Gln Gly Thr				334
100	105		110	
ATA CCT TCC GAA CTT GGA AAT CTG AAG AAT CTC ATC AGC TTG GAT CTG Ile Pro Ser Glu Leu Gly Asn Leu Lys Asn Leu Ile Ser Leu Asp Leu				382
115	120		125	
TAC AAC AAC AAT CTT ACA GGG ATA GTT CCC ACT TCT TTG GGA AAA TTG Tyr Asn Asn Asn Leu Thr Gly Ile Val Pro Thr Ser Leu Gly Lys Leu				430
130	135		140	
AAG TCT CTG GTC TTT TTA CGG CTT AAT GAC AAC CGA TTG ACG GGG CCA Lys Ser Leu Val Phe Leu Arg Leu Asn Asp Asn Arg Leu Thr Gly Pro				478
145	150		155	
ATC CCT AGA GCA CTC ACT GCA ATC CCA AGC CTT AAA GTT GTT GAT GTC Ile Pro Arg Ala Leu Thr Ala Ile Pro Ser Leu Lys Val Val Asp Val				526
160	165	170	175	
TCA AGC AAT GAT TTG TGT GGA ACA ATC CCA ACA AAC GGA CCT TTT GCT Ser Ser Asn Asp Leu Cys Gly Thr Ile Pro Thr Asn Gly Pro Phe Ala				574
180	185		190	
CAC ATT CCT TTA CAG AAC TTT GAG AAC AAC CCG AGG TTG GAG GGA CCG His Ile Pro Leu Gln Asn Phe Glu Asn Asn Pro Arg Leu Glu Gly Pro				622
195	200		205	
GAA TTA CTC GGT CTT GCA AGC TAC GAC ACT AAC TGC ACC TGAAAAAATT Glu Leu Leu Gly Leu Ala Ser Tyr Asp Thr Asn Cys Thr				671
210	215		220	
GGCAAAACCT GAAAATGAAG AATTGGGGGG TGACCTTGTA AGAACACTTC ACCACTTTAT				731
CAAATATCAC ATCTACTATG TAATAAGTAT ATATATGTAG TCCAAAAAAA AAAAAAAA				789

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 220 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

Arg Pro Thr Arg Pro Arg Asn Tyr Arg Trp Glu Leu Phe Ala Ala Ser
1 5 10 15

Leu Ile Leu Thr Leu Ala Leu Ile His Leu Val Glu Ala Asn Ser Glu
20 25 30

Gly Asp Ala Leu Tyr Ala Leu Arg Arg Ser Leu Thr Asp Pro Asp His
35 40 45

Val Leu Gln Ser Trp Asp Pro Thr Leu Val Asn Pro Cys Thr Trp Phe
50 55 60

His Val Thr Cys Asn Gln Asp Asn Arg Val Thr Arg Val Asp Leu Gly
65 70 75 80

Asn Ser Asn Leu Ser Gly His Leu Ala Pro Glu Leu Gly Lys Leu Glu
85 90 95

His Leu Gln Tyr Leu Glu Leu Tyr Lys Asn Asn Ile Gln Gly Thr Ile
100 105 110

Pro Ser Glu Leu Gly Asn Leu Lys Asn Leu Ile Ser Leu Asp Leu Tyr
115 120 125

Asn Asn Asn Leu Thr Gly Ile Val Pro Thr Ser Leu Gly Lys Leu Lys
130 135 140

Ser Leu Val Phe Leu Arg Leu Asn Asp Asn Arg Leu Thr Gly Pro Ile
145 150 155 160

Pro Arg Ala Leu Thr Ala Ile Pro Ser Leu Lys Val Val Asp Val Ser
165 170 175

Ser Asn Asp Leu Cys Gly Thr Ile Pro Thr Asn Gly Pro Phe Ala His
180 185 190

Ile Pro Leu Gln Asn Phe Glu Asn Asn Pro Arg Leu Glu Gly Pro Glu
195 200 205

Leu Leu Gly Leu Ala Ser Tyr Asp Thr Asn Cys Thr
210 215 220

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 894 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..675

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

GGA	CGG	ATT	CAA	GCC	TCC	GAA	GGG	GAC	GCT	CIT	CAC	GCG	CIT	CGC	CGG	48
Gly	Pro	Ile	Gln	Ala	Ser	Glu	Gly	Asp	Ala	Leu	His	Ala	Leu	Arg	Arg	
1		5							10					15		
AGC	TTA	TCA	GAT	CCA	GAC	AAT	GTT	GTT	CAG	AGT	TGG	GAT	CCA	ACT	CIT	96
Ser	Leu	Ser	Asp	Pro	Asp	Asn	Val	Val	Ser	Trp	Asp	Pro	Thr	Leu		
20							25					30				
GTT	AAT	CCT	TGT	ACT	TGG	TTT	CAT	GTC	ACT	TGT	AAT	CAA	CAC	CAT	CAA	144
Val	Asn	Pro	Cys	Thr	Trp	Phe	His	Val	Thr	Cys	Asn	Gln	His	His	Gln	
35							40				45					
GTC	ACT	CGT	CTG	GAT	TTG	GGG	AAT	TCA	AAC	TTA	TCT	GGA	CAT	CTA	GTA	192
Val	Thr	Arg	Leu	Asp	Leu	Gly	Asn	Ser	Asn	Leu	Ser	Gly	His	Leu	Val	
50							55				60					
CCT	GAA	CIT	GGG	AAG	CTT	GAA	CAT	TTA	CAA	TAT	CTG	TAT	GGA	ATC	ATC	240
Pro	Glu	Leu	Gly	Lys	Leu	Glu	His	Leu	Gln	Tyr	Leu	Tyr	Gly	Ile	Ile	
65							70			75			80			
ACT	CTT	TTG	CCT	TTT	GAT	TAT	CTG	AAA	ACA	TTT	ACA	TTA	TCA	GTC	ACA	288
Thr	Leu	Leu	Pro	Phe	Asp	Tyr	Leu	Lys	Thr	Phe	Thr	Leu	Ser	Val	Thr	
85							90				95					
CAT	ATA	ACA	TTT	TGC	TTT	GAG	TCA	TAT	AGT	GAA	CTC	TAC	AAA	AAC	GAG	336
His	Ile	Thr	Phe	Cys	Phe	Glu	Ser	Tyr	Ser	Glu	Leu	Tyr	Lys	Asn	Glu	
100							105				110					
ATT	CAA	GGA	ACT	ATA	CCT	TCT	GAG	CIT	GGA	AAT	CTG	AAG	AGT	CTA	ATC	384
Ile	Gln	Gly	Thr	Ile	Pro	Ser	Glu	Leu	Gly	Asn	Leu	Lys	Ser	Leu	Ile	
115							120				125					
AGT	TTG	GAT	CTG	TAC	AAC	AAC	AAT	CTC	ACC	GGG	AAA	ATC	CCA	TCT	TCT	432
Ser	Leu	Asp	Leu	Tyr	Asn	Asn	Asn	Leu	Thr	Gly	Lys	Ile	Pro	Ser	Ser	
130							135				140					

TTG GGA AAA TIG AAG TCA CTT GTT TTT TTG CGG CTT AAC GAA AAC CGA			480
Leu Gly Lys Leu Lys Ser Leu Val Phe Leu Arg Leu Asn Glu Asn Arg			
145	150	155	160
TTG ACC GGT CCT ATT CCT AGA GAA CTC ACA GTT ATT TCA AGC CTT AAA			528
Leu Thr Gly Pro Ile Pro Arg Glu Leu Thr Val Ile Ser Ser Leu Lys			
165	170	175	
GTT GTT GAT GTC TCA GGG AAT GAT TTG TGT GGA ACA ATT CCA GTA GAA			576
Val Val Asp Val Ser Gly Asn Asp Leu Cys Gly Thr Ile Pro Val Glu			
180	185	190	
GGA CCT TTT GAA CAC ATT CCT ATG CAA AAC TTT GAG AAC AAC CTG AGA			624
Gly Pro Phe Glu His Ile Pro Met Gln Asn Phe Glu Asn Asn Leu Arg			
195	200	205	
TTG GAG GGA CCA GAA CTA CTA GGT CTT GCG AGC TAT GAC ACC AAT TGC			672
Leu Glu Gly Pro Glu Leu Leu Gly Leu Ala Ser Tyr Asp Thr Asn Cys			
210	215	220	
ACT TAAAAAGAAG TTGAAGAACCC TATAAAGAAG AATGTTAGGT GACCTTGAA			725
Thr			
225			
GAACCTCTGTA CCAAGTGTGT GIAAAATCTAT ATAGAGCCCTT GTTTCATGTT ATATATGAAA			785
GCTTTGAGAG ACAGTAACCTT GCAATGTATT GGTATTGGTA GAAAAAGTTG AAATGAGAAT			845
TGCTTTGTA TIGGATTGTG GTTCTTATG TAACTTGAAT TTCTTATTA			894

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 225 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

Gly Pro Ile Gln Ala Ser Glu Gly Asp Ala Leu His Ala Leu Arg Arg
1 5 10 15

Ser Leu Ser Asp Pro Asp Asn Val Val Gln Ser Trp Asp Pro Thr Leu
20 25 30

Val Asn Pro Cys Thr Trp Phe His Val Thr Cys Asn Gln His His Gln
35 40 45

Val Thr Arg Leu Asp Leu Gly Asn Ser Asn Leu Ser Gly His Leu Val
50 55 60

Pro Glu Leu Gly Lys Leu Glu His Leu Gln Tyr Leu Tyr Gly Ile Ile
65 70 75 80

Thr Leu Leu Pro Phe Asp Tyr Leu Lys Thr Phe Thr Leu Ser Val Thr
85 90 95

His Ile Thr Phe Cys Phe Glu Ser Tyr Ser Glu Leu Tyr Lys Asn Glu
100 105 110

Ile Gln Gly Thr Ile Pro Ser Glu Leu Gly Asn Leu Lys Ser Leu Ile
115 120 125

Ser Leu Asp Leu Tyr Asn Asn Leu Thr Gly Lys Ile Pro Ser Ser
130 135 140

Leu Gly Lys Leu Lys Ser Leu Val Phe Leu Arg Leu Asn Glu Asn Arg
145 150 155 160

Leu Thr Gly Pro Ile Pro Arg Glu Leu Thr Val Ile Ser Ser Leu Lys
165 170 175

Val Val Asp Val Ser Gly Asn Asp Leu Cys Gly Thr Ile Pro Val Glu
180 185 190

Gly Pro Phe Glu His Ile Pro Met Gln Asn Phe Glu Asn Asn Leu Arg
195 200 205

Leu Glu Gly Pro Glu Leu Leu Gly Leu Ala Ser Tyr Asp Thr Asn Cys
210 215 220

Thr
225

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1063 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 106..759

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

TCGACCCACG CGTCCGACGA AACCTAAATT TTGCTTCCTC ATCTTGTCAG GAAAATTACT 60

CAAATTCTTA TTAGATTACT CTCTCTTCGA CCTCCGATAG CTCAC ATG GCG TCT 114
Met Ala Ser

1

CGA AAC TAT CGG TGG GAG CTC TTC GCA GCT TCG TTA ATC CTA ACC TTA	162
Arg Asn Tyr Arg Trp Glu Leu Phe Ala Ala Ser Leu Ile Leu Thr Leu	
5 10 15	
GCT TTG ATT CAC CTG GTC GAA GCA AAC TCC GAA GGA GAT GCT CTT TAC	210
Ala Leu Ile His Leu Val Glu Ala Asn Ser Glu Gly Asp Ala Leu Tyr	
20 25 30 35	
GCT CTT CGC CGG AGT TTA ACA GAT CCG GAC CAT GTT CTC CAG AGC TGG	258
Ala Leu Arg Arg Ser Leu Thr Asp Pro Asp His Val Leu Gln Ser Trp	
40 45 50	
GAT CCA ACT CTT GTT AAT CCT TGT ACC TGG TTC CAT GTC ACC TGT AAC	306
Asp Pro Thr Leu Val Asn Pro Cys Thr Trp Phe His Val Thr Cys Asn	
55 60 65	
CAA GAC AAC CGC GTC ACT CGT GTG GAT TTG GGG AAT TCA AAC CTC TCT	354
Gln Asp Asn Arg Val Thr Arg Val Asp Leu Gly Asn Ser Asn Leu Ser	
70 75 80	
GGA CAT CTT GCG CCT GAG CTT GGG AAG CTT GAA CAT TTA CAG TAT CTA	402
Gly His Leu Ala Pro Glu Leu Gly Lys Leu Glu His Leu Gln Tyr Leu	
85 90 95	
GAG CTC TAC AAA AAC AAC ATC CAA GGA ACT ATA CCT TCC GAA CTT GGA	450
Glu Leu Tyr Lys Asn Asn Ile Gln Gly Thr Ile Pro Ser Glu Leu Gly	
100 105 110 115	
AAT CTG AAG AAT CTC ATC AGC TTG GAT CTG TAC AAC AAC AAT CTT ACA	498
Asn Leu Lys Asn Leu Ile Ser Leu Asp Leu Tyr Asn Asn Asn Leu Thr	
120 125 130	
GGG ATA GTT CCC ACT TCT TTG GGA AAA TTG AAG TCT CTG GTC TTT TTA	546
Gly Ile Val Pro Thr Ser Leu Gly Lys Leu Lys Ser Leu Val Phe Leu	
135 140 145	

66 5778 266 208 150

CGG CTT AAT GAC AAC CGA TTG ACC GGG CCA ATC CCT AGA GCA CTC ACT 594
Arg Leu Asn Asp Asn Arg Leu Thr Gly Pro Ile Pro Arg Ala Leu Thr
150 155 160

GCA ATC CCA AGC CTT AAA GTT GTT GAT GTC TCA AGC AAT GAT TTG TGT 642
Ala Ile Pro Ser Leu Lys Val Val Asp Val Ser Ser Asn Asp Leu Cys
165 170 175

GGA ACA ATC CCA ACA AAC GGA CCT TTT GCT CAC ATT CCT TTA CAG AAC 690
Gly Thr Ile Pro Thr Asn Gly Pro Phe Ala His Ile Pro Leu Gln Asn
180 185 190 195

TTT GAG AAC AAC CCG AGG TTG GAG GGA CCG GAA TTA CTC GGT CTT GCA 738
Phe Glu Asn Asn Pro Arg Leu Glu Gly Pro Glu Leu Leu Gly Leu Ala
200 205 210

AGC TAC GAC ACT AAC TGC ACC TGAAAAAATT GGCAAAACCT GAAAATGAAG 789
Ser Tyr Asp Thr Asn Cys Thr
215

AATTGGGGGG TGACCTTGTA AGAACACTTC ACCACTTTAT CAAATATCAC ATCTACTATG 849

TAATAAGTAT ATATATGTAG TCCAAAAAAA AAATGAAGAA TCGAATCAGT AATATCATCT 909

GGTCTCAAATT GAGAACCTTGT AGGTCTGTGT ATGTAAGAATT TCTAAATGCG ACTTTGCGGT 969

ACTGTAATGT TCGGTGTGG GATTCTGAGA AGTAACATTT GTATTGGTAT GGTATCAAGT 1029

TGTTCTGCCT TGTCTGCAAA AAAAAAAAAA AAAA 1063

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 218 amino acids
(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

Met Ala Ser Arg Asn Tyr Arg Trp Glu Leu Phe Ala Ala Ser Leu Ile
1 5 10 15

Leu Thr Leu Ala Leu Ile His Leu Val Glu Ala Asn Ser Glu Gly Asp
20 25 30

Ala Leu Tyr Ala Leu Arg Arg Ser Leu Thr Asp Pro Asp His Val Leu
35 40 45

Gln Ser Trp Asp Pro Thr Leu Val Asn Pro Cys Thr Trp Phe His Val
50 55 60

Thr Cys Asn Gln Asp Asn Arg Val Thr Arg Val Asp Leu Gly Asn Ser
65 70 75 80

Asn Leu Ser Gly His Leu Ala Pro Glu Leu Gly Lys Leu Glu His Leu
85 90 95

Gln Tyr Leu Glu Leu Tyr Lys Asn Asn Ile Gln Gly Thr Ile Pro Ser
100 105 110

Glu Leu Gly Asn Leu Lys Asn Leu Ile Ser Leu Asp Leu Tyr Asn Asn
115 120 125

Asn Leu Thr Gly Ile Val Pro Thr Ser Leu Gly Lys Leu Lys Ser Leu
130 135 140

Val Phe Leu Arg Leu Asn Asp Asn Arg Leu Thr Gly Pro Ile Pro Arg
145 150 155 160

Ala Leu Thr Ala Ile Pro Ser Leu Lys Val Val Asp Val Ser Ser Asn

165

170

175

Asp Leu Cys Gly Thr Ile Pro Thr Asn Gly Pro Phe Ala His Ile Pro

180

185

190

Leu Gln Asn Phe Glu Asn Asn Pro Arg Leu Glu Gly Pro Glu Leu Leu

195

200

205

Gly Leu Ala Ser Tyr Asp Thr Asn Cys Thr

210

215

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2089 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Arabidopsis thaliana

(vii) IMMEDIATE SOURCE:

- (B) CLONE: SERK gene cDNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 195..2069

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

GGATTTTAT TTIATTTTT ACTCTTTGTT TGTTTTAATG CTAATGGGTT TTTAAAAGGG	60
TTATCGAAAA AATGAGTGAG TTGTTGTTGA GGTGTCCTCT GTAAAGTGT AATGGTGGTG	120
ATTTCCGAA GTTACGGTTT TCTGGATCT GAAGAGATCA AATCAAGATT CGAAATTAC	180
CATTGTTGTT TGAA ATG GAG TCG AGT TAT GTG GTG TTT ATC TTA CTT TCA	230
Met Glu Ser Ser Tyr Val Val Phe Ile Leu Leu Ser	
1 5 10	
CTG ATC TTA CTT CCG AAT CAT TCA CTG TGG CTT GCT TCT GCT AAT TTG	278
Leu Ile Leu Leu Pro Asn His Ser Leu Trp Leu Ala Ser Ala Asn Leu	
15 20 25	
GAA GGT GAT GCT TTG CAT ACT TTG AGG GTT ACT CTA GTT GAT CCA AAC	326
Glu Gly Asp Ala Leu His Thr Leu Arg Val Thr Leu Val Asp Pro Asn	
30 35 40	
AAT GTC TTG CAG AGC TGG GAT CCT ACG CTA GTG AAT CCT TGC ACA TGG	374
Asn Val Leu Gln Ser Trp Asp Pro Thr Leu Val Asn Pro Cys Thr Trp	
45 50 55 60	
TTC CAT GTC ACT TGC AAC AAC GAG AAC AGT GTC ATA AGA GTT GAT TTG	422
Phe His Val Thr Cys Asn Asn Glu Asn Ser Val Ile Arg Val Asp Leu	
65 70 75	
GGG AAT GCA GAG TTA TCT GGC CAT TTA GTT CCA GAG CTT GGT GTG CTC	470
Gly Asn Ala Glu Leu Ser Gly His Leu Val Pro Glu Leu Gly Val Leu	
80 85 90	
AAG AAT TTG CAG TAT TTG GAG CTT TAC AGT AAC AAC ATA ACT GGC CCG	518
Lys Asn Leu Gln Tyr Leu Glu Leu Tyr Ser Asn Asn Ile Thr Gly Pro	
95 100 105	
ATT CCT AGT AAT CTT GGA AAT CTG ACA AAC TTA GTG AGT TTG GAT CTT	566

Ile Pro Ser Asn Leu Gly Asn Leu Thr Asn Leu Val Ser Leu Asp Leu
110 115 120

TAC TTA AAC AGC TTC TCC GGT CCT ATT CCG GAA TCA TTG GGA AAG CTT 614
Tyr Leu Asn Ser Phe Ser Gly Pro Ile Pro Glu Ser Leu Gly Lys Leu
125 130 135 140

TCA AAG CTG AGA TTT CTC CGG CTT AAC AAC AAC AGT CTC ACT GGG TCA 662
Ser Lys Leu Arg Phe Leu Arg Leu Asn Asn Ser Leu Thr Gly Ser
145 150 155

ATT CCT ATG TCA CTG ACC AAT ATT ACT ACC CTT CAA GTG TTA GAT CTA 710
Ile Pro Met Ser Leu Thr Asn Ile Thr Thr Leu Gln Val Leu Asp Leu
160 165 170

TCA AAT AAC AGA CTC TCT GGT TCA GTT CCT GAC AAT GGC TCC TTC TCA 758
Ser Asn Asn Arg Leu Ser Gly Ser Val Pro Asp Asn Gly Ser Phe Ser
175 180 185

CTC TTC ACA CCC ATC AGT TTT GCT AAT AAC TTA GAC CTA TGT GGA CCT 806
Leu Phe Thr Pro Ile Ser Phe Ala Asn Asn Leu Asp Leu Cys Gly Pro
190 195 200

GTT ACA AGT CAC CCA TGT CCT GGA TCT CCC CCG TTT TCT CCT CCA CCA 854
Val Thr Ser His Pro Cys Pro Gly Ser Pro Pro Phe Ser Pro Pro Pro
205 210 215 220

CCT TTT ATT CAA CCT CCC CCA GTT CCT ACC CCG AGT GGG TAT GGT ATA 902
Pro Phe Ile Gln Pro Pro Pro Val Ser Thr Pro Ser Gly Tyr Gly Ile
225 230 235

ACT GGA GCA ATA GCT GGT GGA GTT GCT GCA GGT GCT GCT TTG CCC TTT 950
Thr Gly Ala Ile Ala Gly Gly Val Ala Ala Gly Ala Ala Leu Pro Phe
240 245 250

GCT GCT CCT GCA ATA GCC TTT GCT TGG TGG CGA CGA AGA AGC CCA CTA 998
Ala Ala Pro Ala Ile Ala Phe Ala Trp Trp Arg Arg Arg Ser Pro Leu

255	260	265	
GAT ATT TTC TIC GAT GTC CCT GCC GAA GAA GAT CCA GAA GTT CAT CTG Asp Ile Phe Phe Asp Val Pro Ala Glu Glu Asp Pro Glu Val His Leu			1046
270	275	280	
GGA CAG CTC AAG AGG TTT TCT TTG CGG GAG CTA CAA GTG GCG AGT GAT Gly Gln Leu Lys Arg Phe Ser Leu Arg Glu Leu Gln Val Ala Ser Asp			1094
285	290	295	300
GGG TTT AGT AAC AAG AAC ATT TTG GGC AGA GGT GGG TTT GGG AAA GTC Gly Phe Ser Asn Lys Asn Ile Leu Gly Arg Gly Phe Gly Lys Val			1142
305	310	315	
TAC AAG GGA CGC TTG GCA GAC GGA ACT CTT GTT GCT GTC AAG AGA CTG Tyr Lys Gly Arg Leu Ala Asp Gly Thr Leu Val Ala Val Lys Arg Leu			1190
320	325	330	
AAG GAA GAG CGA ACT CCA GGT GGA GAG CTC CAG TTT CAA ACA GAA GTA Lys Glu Glu Arg Thr Pro Gly Gly Glu Leu Gln Phe Gln Thr Glu Val			1238
335	340	345	
GAG ATG ATA AGT ATG GCA GTT CAT CGA AAC CTG TTG AGA TTA CGA GGT Glu Met Ile Ser Met Ala Val His Arg Asn Leu Leu Arg Leu Arg Gly			1286
350	355	360	
TTC TGT ATG ACA CCG ACC GAG AGA TTG CTT GTG TAT CCT TAC ATG GCC Phe Cys Met Thr Pro Thr Glu Arg Leu Leu Val Tyr Pro Tyr Met Ala			1334
365	370	375	380
AAT GGA AGT GTT GCT TCG TGT CTC AGA GAG AGG CCA CCG TCA CAA CCT Asn Gly Ser Val Ala Ser Cys Leu Arg Glu Arg Pro Pro Ser Gln Pro			1382
385	390	395	
CCG CTT GAT TGG CCA ACG CGG AAG AGA ATC GCG CTA GGC TCA GCT CGA Pro Leu Asp Trp Pro Thr Arg Lys Arg Ile Ala Leu Gly Ser Ala Arg			1430
400	405	410	

BIOCHEMICAL DATA

GGT TTG TCT TAC CTA CAT GAT CAC TGC GAT CGG AAG ATC ATT CAC CGT Gly Leu Ser Tyr Leu His Asp His Cys Asp Pro Lys Ile Ile His Arg 415 420 425	1478
GAC GTA AAA GCA GCA AAC ATC CTC TTA GAC GAA GAA TTC GAA GCG GTT Asp Val Lys Ala Ala Asn Ile Leu Leu Asp Glu Glu Phe Glu Ala Val 430 435 440	1526
GTT GGA GAT TTC GGG TTG GCA AAG CTT ATG GAC TAT AAA GAC ACT CAC Val Gly Asp Phe Gly Leu Ala Lys Leu Met Asp Tyr Lys Asp Thr His 445 450 455 460	1574
GTG ACA ACA GCA GTC CGT GGC ACC ATC GGT CAC ATC GCT CCA GAA TAT Val Thr Thr Ala Val Arg Gly Thr Ile Gly His Ile Ala Pro Glu Tyr 465 470 475	1622
CTC TCA ACC GGA AAA TCT TCA GAG AAA ACC GAC GTT TTC GGA TAC GGA Leu Ser Thr Gly Lys Ser Ser Glu Lys Thr Asp Val Phe Gly Tyr Gly 480 485 490	1670
ATC ATG CTT CTA GAA CTA ATC ACA GGA CAA AGA GCT TTC GAT CTC GCT Ile Met Leu Leu Glu Leu Ile Thr Gly Gln Arg Ala Phe Asp Leu Ala 495 500 505	1718
CGG CTA GCT AAC GAC GAC GTC ATG TTA CTT GAC TGG GTG AAA GGA Arg Leu Ala Asn Asp Asp Val Met Leu Leu Asp Trp Val Lys Gly 510 515 520	1766
TTG TTG AAG GAG AAG CTA GAG ATG TTA GTG GAT CCA GAT CTT CAA Leu Leu Lys Glu Lys Lys Leu Glu Met Leu Val Asp Pro Asp Leu Gln 525 530 535 540	1814
ACA AAC TAC GAG GAG AGA GAA CAA GTG ATA CAA GTG GCG TTG Thr Asn Tyr Glu Glu Arg Glu Leu Glu Gln Val Ile Gln Val Ala Leu 545 550 555	1862

CTA TGC ACG CAA GGA TCA CCA ATG GAA AGA CCA AAG ATG TCT GAA GTT 1910
Leu Cys Thr Gln Gly Ser Pro Met Glu Arg Pro Lys Met Ser Glu Val
560 565 570

GTA AGG ATG CTG GAA GGA GAT GGG CTT GCG GAG AAA TGG GAC GAA TGG 1958
Val Arg Met Leu Glu Gly Asp Gly Leu Ala Glu Lys Trp Asp Glu Trp
575 580 585

CAA AAA GTT GAG ATT TTG AGG GAA GAG ATT GAT TTG AGT CCT AAT CCT 2006
Gln Lys Val Glu Ile Leu Arg Glu Glu Ile Asp Leu Ser Pro Asn Pro
590 595 600

AAC TCT GAT TGG ATT CTT GAT TCT ACT TAC AAT TTG CAC GCC GTT GAG 2054
Asn Ser Asp Trp Ile Leu Asp Ser Thr Tyr Asn Leu His Ala Val Glu
605 610 615 620

TTA TCT GGT CCA AGG TAAAAAAAAA AAAAAAAA 2089
Leu Ser Gly Pro Arg
625

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 625 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

Met Glu Ser Ser Tyr Val Val Phe Ile Leu Leu Ser Leu Ile Leu Leu
1 5 10 15

Pro Asn His Ser Leu Trp Leu Ala Ser Ala Asn Leu Glu Gly Asp Ala
20 25 30

Leu His Thr Leu Arg Val Thr Leu Val Asp Pro Asn Asn Val Leu Gln
35 40 45

Ser Trp Asp Pro Thr Leu Val Asn Pro Cys Thr Trp Phe His Val Thr
50 55 60

Cys Asn Asn Glu Asn Ser Val Ile Arg Val Asp Leu Gly Asn Ala Glu
65 70 75 80

Leu Ser Gly His Leu Val Pro Glu Leu Gly Val Leu Lys Asn Leu Gln
85 90 95

Tyr Leu Glu Leu Tyr Ser Asn Asn Ile Thr Gly Pro Ile Pro Ser Asn
100 105 110

Leu Gly Asn Leu Thr Asn Leu Val Ser Leu Asp Leu Tyr Leu Asn Ser
115 120 125

Phe Ser Gly Pro Ile Pro Glu Ser Leu Gly Lys Leu Ser Lys Leu Arg
130 135 140

Phe Leu Arg Leu Asn Asn Asn Ser Leu Thr Gly Ser Ile Pro Met Ser
145 150 155 160

Leu Thr Asn Ile Thr Thr Leu Gln Val Leu Asp Leu Ser Asn Asn Arg
165 170 175

Leu Ser Gly Ser Val Pro Asp Asn Gly Ser Phe Ser Leu Phe Thr Pro
180 185 190

Ile Ser Phe Ala Asn Asn Leu Asp Leu Cys Gly Pro Val Thr Ser His
195 200 205

Pro Cys Pro Gly Ser Pro Pro Phe Ser Pro Pro Pro Pro Phe Ile Gln
210 215 220

Pro Pro Pro Val Ser Thr Pro Ser Gly Tyr Gly Ile Thr Gly Ala Ile
225 230 235 240

Ala Gly Gly Val Ala Ala Gly Ala Ala Leu Pro Phe Ala Ala Pro Ala
245 250 255

Ile Ala Phe Ala Trp Trp Arg Arg Arg Ser Pro Leu Asp Ile Phe Phe
260 265 270

Asp Val Pro Ala Glu Glu Asp Pro Glu Val His Leu Gly Gln Leu Lys
275 280 285

Arg Phe Ser Leu Arg Glu Leu Gln Val Ala Ser Asp Gly Phe Ser Asn
290 295 300

Lys Asn Ile Leu Gly Arg Gly Gly Phe Gly Lys Val Tyr Lys Gly Arg
305 310 315 320

Leu Ala Asp Gly Thr Leu Val Ala Val Lys Arg Leu Lys Glu Glu Arg
325 330 335

Thr Pro Gly Gly Glu Leu Gln Phe Gln Thr Glu Val Glu Met Ile Ser
340 345 350

Met Ala Val His Arg Asn Leu Leu Arg Leu Arg Gly Phe Cys Met Thr
355 360 365

Pro Thr Glu Arg Leu Leu Val Tyr Pro Tyr Met Ala Asn Gly Ser Val
370 375 380

Ala Ser Cys Leu Arg Glu Arg Pro Pro Ser Gln Pro Pro Leu Asp Trp
385 390 395 400

Pro Thr Arg Lys Arg Ile Ala Leu Gly Ser Ala Arg Gly Leu Ser Tyr
405 410 415

Leu His Asp His Cys Asp Pro Lys Ile Ile His Arg Asp Val Lys Ala

420

425

430

Ala Asn Ile Leu Leu Asp Glu Glu Phe Glu Ala Val Val Gly Asp Phe

435

440

445

Gly Leu Ala Lys Leu Met Asp Tyr Lys Asp Thr His Val Thr Thr Ala

450

455

460

Val Arg Gly Thr Ile Gly His Ile Ala Pro Glu Tyr Leu Ser Thr Gly

465

470

475

480

Lys Ser Ser Glu Lys Thr Asp Val Phe Gly Tyr Gly Ile Met Leu Leu

485

490

495

Glu Leu Ile Thr Gly Gln Arg Ala Phe Asp Leu Ala Arg Leu Ala Asn

500

505

510

Asp Asp Asp Val Met Leu Leu Asp Trp Val Lys Gly Leu Leu Lys Glu

515

520

525

Lys Lys Leu Glu Met Leu Val Asp Pro Asp Leu Gln Thr Asn Tyr Glu

530

535

540

Glu Arg Glu Leu Glu Gln Val Ile Gln Val Ala Leu Leu Cys Thr Gln

545

550

555

560

Gly Ser Pro Met Glu Arg Pro Lys Met Ser Glu Val Val Arg Met Leu

565

570

575

Glu Gly Asp Gly Leu Ala Glu Lys Trp Asp Glu Trp Gln Lys Val Glu

580

585

590

Ile Leu Arg Glu Glu Ile Asp Leu Ser Pro Asn Pro Asn Ser Asp Trp

595

600

605

Ile Leu Asp Ser Thr Tyr Asn Leu His Ala Val Glu Leu Ser Gly Pro

610

615

620

Arg
625